

Introduction to PAUP* – Barber Lab (PB & EC)

- 1) PAUP* and Modeltest
- 2) PAUP*'s Diverse Tree Search Methods
- 3) Bootstrapping
- 4) Kishino-Hasegawa Test

PAUP* (Phylogenetic Analysis Using Parsimony...*and other methods) is one of the most highly used phylogenetic software programs. The new version does distance, parsimony and maximum likelihood methods. We will be using PAUP* in conjunction with MODELTEST, a program that evaluates the best molecular model of evolution for your data.

1) PAUP* and Modeltest

1. Open PAUP*, and it will ask you which file to open. Select your file, and the *Edit* radio button. The PAUP editor is better than any text editor for dealing with the nuts and bolts of a nexus file. As a quick introduction to nexus format you will see that the file is divided into several blocks of data and commands. Each block is delimited by BEGIN and END. The first block is the DATA block, in which you should see your data interleaved or uninterleaved, and the sequence length and number of taxa denoted.
2. You are now going to insert a new block of code, written for Modeltest, that will give PAUP a set of specific commands:
 - a. *File* → *Open* → *Browse* to Modeltest folder → Batch folder and open ModelblockPAUPb10 and *Edit* radiobutton.
 - b. Use *Select All* to copy the text of this file and paste the entire text at the bottom of your file (you can remove the #Nexus though)
 - c. *Save* the file back in its original directory.
3. You will now ask PAUP *to actually read your file, along with the new commands.
 - a. *File* → *Open* → *Select* your file and the *Execute* radio button.
 - b. A screen should come up with a bunch of text that will change frequently. PAUP* has built a NJ tree and is calculating the likelihood score of each tree given the data and 1 of 56 different models of evolution. It may take 5-10 minutes to complete.
 - c. When run is complete, a file called *model.scores* will be created in the same folder that the original file was in.
4. Now use Modeltest to evaluate the likelihood scores generated by PAUP (which you can see if you open *model.scores*, or better yet, read the logfile that was also generated)to find the best model of evolution.
 - a. Open modeltest.mac (in bin folder in Modeltest folder)
 - b. Select the input file (*model.scores*) and the output destination and filename.
 - c. The program will run in a fraction of a second and the output file will be created.
5. In PAUP*, open your file and the output file that you created from ModelTest. The later will have a block of likelihood scores and a list of hierarchical of likelihood ratio test results, as well as results from a similar test of the Akaike information criterion. Before the start of the Akaike section, modeltest has written the model that it selected based on the hierarchical test, along with associated parameters. Beneath that is a block of instructions for PAUP* to set this model as the model for a likelihood search.
6. *Delete* the modelblock from your file.
7. *Copy* the PAUP block from the modeltest output starting from “[!Likelihood settings...” to “end;” (be sure to include [] and ;)
8. *Paste* the modelblock to the very bottom of the Allsilversides file and *Save*.
9. *File* → *Execute* your file. Note that Likelihood settings are set to the chosen model.

2) PAUP*'s Diverse Tree Search Methods

1. First, define an appropriate outgroup.
 - a. *Data* → *Define Outgroup* – move one of the taxa into the box.
 - b. *Options* → *Rooting* – select *Make Ingroup Monophyletic*
2. Create a neighbor joining tree based on the p-distances between sequences. (NB – Neighbor-Joining or UPGMA trees use clustering algorithms to join closely related taxa. They do not search over multiple trees to optimize a criterion like parsimony or likelihood.)
 - a. *Analysis* → *Distance* – This will set the optimality criterion to Distance
 - b. *Analysis* → *Distance setting*
 - i. Select the *DNA/RNA distances* button, and *Uncorrected P-distance*
 - c. *Analysis* → *Neighborjoining/UPGMA* - select *Neighborjoining*
 - d. *Trees* → *Tree scores* – write down this value
 - e. *Trees* → *Print NJ Tree*
 - i. Change *Plot type* to Phylogram
 - ii. Hit *Preview*.
 - iii. Save the tree as a PICT file
3. Now create a Parsimony tree:
 - a. *Analysis* → *Parsimony* - This will set the optimality criterion to Parsimony
 - b. *Analysis* → *Heuristic search*
 - i. Select *Stepwise-addition options*
 - ii. Choose *Random*, with 10 reps – this ensures the best search of treespace.
 - iii. Hit search
 - c. If you get more than one best tree from this search, you may want to make a consensus tree:
 - i. *Trees* → *Compute Consensus*
 - d. *Trees* → *Tree scores* – write down this value
 - e. *Trees* → *Print Trees...* (or *print consensus*)
 - i. Change *Plot type* to Phylogram
 - ii. Hit *Preview*.
 - iii. Save the tree as a PICT file
4. Finally, create a Likelihood tree:
 - a. *Analysis* → *Likelihood* - This will set the optimality criterion to Likelihood.
 - b. *Analysis* → *Likelihood settings* – You'll notice that the settings have been set by Modeltest
 - c. *Analysis* → *Heuristic search*
 - i. Select *Stepwise-addition options*
 - ii. Choose *Random*, with 10 reps – this ensures the best search of treespace.
 - iii. Hit search
 - d. *Trees* → *Tree scores* – write down this value
 - e. *Trees* → *Print Trees...* (or *print consensus*)
 - i. Change *Plot type* to Phylogram
 - ii. Hit *Preview*.
 - iii. Save the tree as a PICT file

NB – please upload your PICT files to the website, with your name and type of tree in the filename.

Question: Based on your treescores and topologies, which of the three trees best reflects the true history? Is it possible to compare the treescores from the different methods? Why or why not?

3) Bootstrapping

1. *Execute* your data file
2. Define your outgroup, as done before.
3. Set optimality criterion to *parsimony*
4. *Analysis* → *bootstrap/jackknife*
5. Select *bootstrap*, and set it for 100 Replicates. Retain all groups with Frequencies greater than 60%.
6. Hit Continue and Search
7. *Trees* → *print bootstrap consensus*
8. Select *Rectangular Cladogram*. Select *Show Group Frequencies*
9. Print the bootstrap consensus tree

Exercise: Complete and print 2 bootstrap consensus trees. In one tree, keep groups with frequencies greater than 60%. In the second, keep groups with frequency greater than 80%.

4) Kishino-Hasegawa Test

1. In MacClade, open your data file, and construct a constraint tree that reflects an alternative hypothesis of evolutionary history.
 - a. *Display* → *Go to tree window*
 - b. Select *Display random bush*
 - c. Dragging individual branches, construct a topology that is different from the one chosen by PAUP*. (Example – with population data, construct a hypothesis of reciprocal monophyly...with species level data, construct a hypothesis of genus polytomy - no phylogenetic signal – you can use the *collapse branches* tool)
 - d. *Tree* → *Store Tree* → Save your tree with an appropriate name.
 - e. Save your file.
2. Now use PAUP* to search for the best tree with your constraint:
 - a. Open your file in PAUP* and hit *execute*.
 - b. Define your outgroup
 - c. Set optimality criterion to *parsimony*
 - d. *Analysis* → *Load constraint*
 - e. Open your constraint
 - f. *Analysis* → *Heuristic Search*
 - g. In *General Search Options*, click *Enforce Topological Constraints*, keeping trees that *Are Compatible* with constraints
 - h. Hit *Search*
 - i. Save tree to file *HKTest.tre*
3. And find the best tree without constraints:
 - a. *Analysis* → *Heuristic Search*
 - b. In *General Search Options*, unclick *Enforce Topological Constraints*, so that the heuristic search will no longer enforce the constraint tree.
 - c. Hit *Search*
 - d. Save trees to the same file *HkTest*. When asked, say *Append File*. You should now have two trees in the *HkTest* treefile
4. Now compare your two trees:
 - a. Open the *Hk Test* treefile in the PAUP* editor

- b. Rather than adding all the trees together in one tree block, APPENDING the tree file merges two independent files together. You must now edit the file to create a single tree block. First, label the tree(s) that you got with the constraint so that you know what is what. Then, copy one tree block (e.g. "tree PAUP_1 =") into the other so that there is only one treeblock
5. *Trees* → *Get Trees from file*
6. Open *HkTest* treefile and press *Get Trees*. Two trees should be imported into PAUP.
7. *Trees* → *Trees Scores*, and click *Likelihood*
8. Click *Topology Tests*
9. Click *Kishino-Hasegawa Test*, *Normal distribution*, and *one-tailed* and hit OK

Exercise:

Describe how you would use the Kishino-Hasegawa test to test the hypothesis that there has been no interbreeding between two populations for a long enough time to make them reciprocally monophyletic.